THE RELATION OF LIBRARY CATALOGS TO ABSTRACTING AND INDEXING SERVICES

By

Frank B. Rogers
Director, National Library of Medicine

It is probable that the title of this paper would make more sense if it were inverted, or at least, if it were modified to read "The Interrelationship of Library Catalogs and Abstracting and Indexing Services." Throughout this paper I will use the brief form "indexes" to mean indexing and abstracting services.

The great flowering of the periodical form of publication occurred in the nineteenth century, and following closely came the rise of collective indexes to the periodical literature, themselves in periodical form. From the beginning, the traditional librarian had been impressed with the primacy of the book in its traditional format, and he paid scant heed to the flimsy and presumably ephemeral periodicals. With a few exceptions, notably Poole, periodical indexes were constructed and published outside the main stream of librarianship. The chief indexes were sponsored by

---

Presented on August 7, 1963, at the 28th Annual Conference of the University of Chicago Graduate Library School, "Library Catalogs: Changing Dimensions."

Arch.
Z
695.9
R725r
1963
c.1

NATIONAL LIBRARY OF MEDICINE
BETHESDA 14, MD.

scientific societies and organizations other than libraries; this separatist tradition has been carried down to the present day.

One of the problems which the indexes had to face up to earlier than did the libraries was the problem of volume of literature. As objects requiring bibliographical processing, the number of periodical articles stands to the number of books roughly in the ratio 20 to 1. The large indexes of today regularly catalog more items in their special fields than do the largest research libraries of universal scope.

It is a noteworthy fact that indexes were and are typically issued in codex form, whereas library catalogs have been typically of the card variety. It is interesting to note, also, that as the problem of volume presses libraries more and more severely, the interest in library catalogs in book form continues to mount.

One of the most significant features of publication of indexes in codex form is that this means that they are time segmented. They are themselves periodicals. They may be cumulated, to be sure, but cumulation has its economic limits, and time segmentation remains. This in itself gives the index tremendous advantages in flexibility of arrangement

and organization. It enables the index to adapt itself more readily to changing ideas and changing interests. The index does not carry constantly on its back, as does the library card catalog, the awful burden of the past. There are trade-offs here, of course; the index cumulation published yesterday cannot show us the edition published today, in relation to all the previously published editions, but the card catalog can. The point is that in the context of the increasing volume of the literature, this sort of trade-off begins to appear more and more attractive. Time segmentation has some very large advantages. Perhaps we are seeing the beginning of a trend in the publication, by various libraries, of catalogs on a century basis.

Another important feature of the index, as contrasted with the library catalog, is its overwhelming emphasis on subject cataloging. In the periodical index, subject cataloging has gained the ascendancy over descriptive cataloging. The descriptive cataloging has become just a very complicated address number. The author's name has become, in a sense, just another tag which has a bearing on the validity or credibility of the subject content of the article. It is possible to interpret this, in a philosophical sense,

as a phenomenon connected with the emergence of mass-man and mass-society. We may choose to view the phenomenon with some regrets, but we cannot afford to fail to recognize its existence and its implications.

These differences of approach and emphasis lead to some significant differences in practice. Differences in descriptive cataloging practice are striking.

One of the largest problems of descriptive cataloging of books inheres in the matter of corporate author entry; this is a problem only rarely encountered in the descriptive cataloging of periodical articles.

At the same time, no large periodical index does any authority work on the establishment of personal author names. This is not to say that the periodical index is not concerned with problems such as Spanish surname order, or transposition of some foreign names from the genitive to the nominative case. But in the periodical index such problems are handled by rule, and devil take the hindmost; initials are just initials, and often first names may be converted to initials, rather than the other way around. It is inevitable that such practices occasionally do violence to the traditional first principle of descriptive cataloging, that of bringing together under

a standard form of name all of an author's works. The fact that the index is published in codex form, chronologically segmented, inevitably means also that the traditional second principle of descriptive cataloging, that of bringing together all editions of an author's work, is also occasionally violated.

In the area of subject cataloging, interestingly enough, the differences are more apparent than real. The librarian has been unhappily misled by the assumption, long held in the profession, that the subject indexing of periodical articles is one thing, but the subject cataloging of books is quite another. This false assumption rested on a comparison with the indexing of individual books, each one made on an ad hoc basis, and not on a comparison with periodical indexes, each itself a periodical with a continuing life and a continuing need for internal consistency.

If we stop for a moment, and try to paint a broad brush picture of the situation as it existed at the mid-point of the twentieth century, we would see two enormous bibliographical efforts proceeding side by side, the products of separate institutional forms, with separate organizational philosophies. The wonder is that the catalog and the index, at the consumer end of the line so universally recognized as being complementary and interlocking and mutually indispensable,

should at the production end of the line have been presumed to have no relevance to each other.

Again at the mid-point of the twentieth century, we might try to make a snapshot of another emerging phenomenon. It is apparent that a technological revolution is under way. And it is recognizable that to describe it as a revolution is not indulging in hyperbole. And one of the machines that bursts upon the scene is the computer, as significant for society as was the appearance of the steam engine in its day.

At the same time, in industry and technology predominantly, the library as an institutional form is being reinvented and renamed, and is being manned in most cases by people without formal library education and experience. The reasons for this are varied: they revolve partly around the emergence of yet a new form of publication, the project report; partly around the security requirements of the age of the cold war; partly around the complexities of the increasingly interdisciplinary, mission-oriented character of research and development work. Feeding on, and being fed by, the affluence around it, the new technology spirals into an ever higher pitch of publication, and the bibliographical main-stream acquires great additional masses of

material to assimilate.

The new breed of documentalist, often not aware that he is functioning as a librarian, and not inhibited by the folklore of the profession, begins to recognize that the powers of comparison inherent in the computer, and its capability for performing long sequences of work, branching in the middle of the process in one direction or another, according to this or that result achieved during the early stage of the process, are potentially great bibliographic tools. He sees that the greater-than and less-than comparisons are the essence of sorting and file arrangement and rearrangement. More importantly, he senses in the equal-to and not-equal-to comparisons the basis for a new and powerful approach to subject bibliography. He experiments and he fumbles and he experiments some more, and eventually he comes up with a notion of profound importance, which becomes generally referred to as coordinate indexing.

Here I pause to say that, as most of you realize, that isn't the way it actually happened. There are many reservations, temporally and otherwise, which one would have to make to that account. But I would maintain that that <u>is</u> the way that it happened, philosophically, ideationally, and in non-historical but logical sequence.

I say the concept of coordinate indexing was profound. Because I am a librarian, speaking to an audience of librarians, it is necessary for me to assure you that I do not say this in irony, but in truth.

The concept of coordinate indexing is very simple, and is very closely related to traditional notions of subject cataloging. But just the slight twist makes a powerful difference. It is like the difference between Euclidean and Riemannian geometry. Euclidean geometry is still adequate for the construction of buildings, but it is not adequate for the exploration of outer space.

Moreover, coordinate indexing is pre-eminently the machine mode, the computer mode, the mode in which the great power of the computer can best be realized in subject bibliography.

There is one important additional fact to be noted in this connection. These early uses of the computer for subject control of documents, and indeed present usage of the computer for the subject control of documents, as exemplified in the chemical and petroleum industries, for example, are not conceived of in terms of library systems but in terms of individual libraries. They are not substitutes for the

published catalog which gets widespread dissemination; they are substitutes for the individual card subject catalog, for the individual enterprise. The terms which carry the subject concepts are hidden in the bowels of the machine; the answers emerge in fragments, made explicit only at the time of search.

Now these people who came up with computer usage and coordinate indexing were, as we have seen, para-library types, rather than librarians. We have also seen that the people who compose and produce indexes are also para-library types. It is perhaps natural that they should get together before either of them was embraced by the librarian.

The index producer had a different need. He was producing a printed book. The technics of printing were an ever-present source of concern to him. And he eventually saw that the computer was not only a manipulatory device, but that it had major printing possibilities.

And so he began to experiment with these possibilities. We may say, philosophically speaking once more, and not in strict time sequence, that the first fruit of his efforts is the KWIC, keyword in context index. The KWIC index is quick; it is cheap, and it is certainly effective, especially as a current awareness alerting device. It may also be described

as dirty. It was dependent upon the vagaries of uncontrolled natural language, to which are added the pitfalls of translation. It did not have the convenience of the unit-record style of entry, where all necessary information is to be found in one place. These defects cause its capabilities as a retrospective searching tool to suffer considerably.

What has been happening during the last few years is that an attempt is being made at synthesis of the two capabilities and the two objectives. What has been going on at NASA, at the National Library of Medicine, at Chemical Abstracts Service, and elsewhere, is the attempt to produce, within a single complex system, a bibliographic store within the computer which will respond beautifully to the one-shot demand search of great complexity, and at the same time will periodically provide a print-out of the store, to be disseminated in book form.

The trouble is that there are baffling antinomies inherent in such a dual system. If we wish to use machines for the publication of indexes, we must worry very much about the expense of posting and printing each unit record under each of a very large number of subject rubrics. Worse, we must be aware that in the manual search of a tray of

catalog cards, or of a printed index of the ordinary type, or of a printed subject catalog, the user does not proceed by looking at all the entries under all the terms which define the subject of his inquiry,and then trying to find out which citations are common to all the rubrics. That is what the machine does, but it is not how the human being operates.

We are justified, however, in trying to devise systems which will embrace both areas, that is, will respond readily to one-shot demand searches of great complexity, and will also output printed copy suitable for publication. These requirements are disparate in character; it is something like the problem of trying to build a combination lawnmower and electric shaver which is at once economical, easy to operate in either mode, and effective in each. We are justified in making the attempt because in both cases we are operating on the same store of bibliographical materials. And it is the initial winning of this store, and the digitalizing of it, which is the main expense in either mode. There is good reason for trying to make a single system serve both ends. Indeed, it may be that economically viable machine systems are feasible only when they embrace these dual characteristics.

The type of compromise demanded lies in the area of the type of subject terms to be used in the system. And the first thing needed is the realization of the fact that there are indeed different types of terms available. Nothing is more discouraging, in an otherwise earnest discussion of subject cataloging, than to hear a statement such as "subject headings, descriptors, uniterms, keywords, or whatever you call them," as if all were synonymous. It makes all the difference what you call them, and what they are.

The basic choice is between the use of a controlled language, or an uncontrolled language. If we choose to use an uncontrolled language, then we are talking about uniterms or keywords. I will simply assert here, without pursuing the argument further, that in the case of massive catalogs or massive indexes, the choice of an uncontrolled language for subject analysis is a poor one.

We will be using a controlled language, of which there are two main varieties, the subject heading and the descriptor.

The subject heading is typically a noun plus modifier, or a phrase, which in itself involves a pre-coordination of terms. The unwritten premise of traditional library subject cataloging has been the attempt to provide a single most-specific subject heading which fits the work or the article

as a whole, a single subject heading which as a short phrase is somehow a compressed encapsulation of the precise subject being denoted. The subject-heading is a label which says, in effect, "This is it."

In contrast, the descriptor is a pointer which says "There it is," and the intersect of two or more pointers will define a very specific subject. The descriptor is more elemental, basic, broader in meaning. There is, in descriptor systems, more concern with consistency of level of the terminology chosen. There is much more concern with eliminating what J. E. Holmstrom has so aptly called "the quasi-penumbral synonym."

The number of subject headings in a given system will be very much larger than the number of descriptors required for the same system. More importantly, the relationships between and among subject headings are very much more complicated and difficult to control than are the relationships between and among descriptors.

The descriptor is the language of coordinate indexing, and of pure machine retrieval, without consideration of publication.

The subject heading is the language of publication of

subject bibliographies made up of unit records. It is more economical of printing space, but it has less retrieval power, because it is limited to logical sums.

The compromise needed is therefore a subject cataloging and indexing language which lies somewhere between subject headings and descriptors.

In both systems, what we are doing is to use words of the natural language in a somewhat stilted and stereotyped way, which is desirable. The nature of language provides many traps for the system builder, however; it is the nature of language to provide new single words which are in themselves a pre-coordination of other terms. The suffixes and prefixes in common usage provide many examples. If we have in our subject system the term APPENDIX, from the category of anatomical names, and the term INFLAMMATION, from the category of pathological processes, how can we avoid, even in a descriptor system, the use of the term APPENDICITIS, from the category of disease nomenclature? The answer is that we cannot avoid it and still remain within the bounds of reason and reasonable usage. There is a very large gray area here which makes compromise almost a relief, rather than a burden.

Nevertheless, and in spite of the gray area, it is a good thing to have some conceptual framework on which planning may be based, some hypothesis or theory against which we may test what we are doing. The new Index Medicus which will emerge from the National Library of Medicine computer sometime next spring will illustrate how well or how poorly these points have been calculated. It will also furnish a concrete example of the relationship between catalogs and indexes, for it will list both books and journal articles, interspersed in the same sections. In a way, with some new twists, it will be a return to John Shaw Billings' Index-Catalogue, and that hyphenated term is significant of much.

In summary, the growth and development of abstracting and indexing services have several important implications for cataloging. They demonstrate dramatically the heightening of problems which occurs when the volume of bibliographic items is tremendously increased. They demonstrate some virtues of the book format. They accentuate the primary importance of subject cataloging. They call into question the fundamental principles of descriptive cataloging as it is now commonly practiced. They serve as pathfinders in the accelerating drive toward finding suitable machine

solutions to the most important library problem which we face today -- improved bibliographic access to the literature.

I would like to close by stating some articles of my personal faith, and leaving you with a question.

I believe that it is established that in massive bibliographic structures the computer, using a coordinate indexing system with a descriptor language, is the best available answer to the problem of providing really adequate responses to one-shot demand searches of great complexity.

I believe that it will be demonstrated that the economic viability of large machine bibliographic systems, in which the bibliographic store has been mined at great effort and cost, depends on the added capability of printing out periodic catalogs and indexes. These printed catalogs and indexes provide the most efficient medium for search from the standpoint of widespread accessibility, coupled with rapidity of response, in the great majority of bibliographic inquiries. Although they cannot yield the same intensity and depth of response that the rarer complex search demands, they are adequately responsive to that much larger number of questions which are yet of great social utility and importance.

I believe we must compromise to achieve these dual objectives, and I think it is quite probable that the nature of the compromise will be along the lines I have indicated, that is, conventional library subject cataloging practice must be modified in the direction of the descriptor end of the continuum.

There remains the enormous question of the effect of subject specialization. It is a fact of great importance that most of the major abstracting and indexing services are limited to one subject field, usually supradisciplinary in character. Systems of coordinate indexing are clearly applicable to each such individual field. Can they be applied universally, across the board, to universal collections, universal in scope? The collections of the Library of Congress are 15 to 20 times as broad as those of the National Library of Medicine, but it seems to me that for subject cataloging systems of equal responsiveness and value, the Library of Congress has a problem far more than 20 times as complicated and as expensive to solve as does the National Library of Medicine.

For myself, I tend to be skeptical of our ability to convert the descriptor systems of various fields into one

big descriptor system of universal scope. It is meaningless to assign a certain percentage of compatibility between the systems of the National Library of Medicine and the Library of Congress, for example, merely because both systems contain the term GEOLOGY and both contain the term MEDICINE. We seem to be faced with a situation in which either present subject cataloging in general collections must be intensified and expanded by an order of magnitude, or the system of separate subject controls in separate special areas must be widened and rationalized. There is some reason to believe that the dual-track system now utilized, with ankle-deep subject cataloging of general collections, and knee-deep subject cataloging of special collections, is the only really feasible arrangement. This implies a continuation of the polycentrism which now characterizes our bibliographic world.

* * * * *